Talanta 91 (2012) 72–76

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Direct matrix assisted laser desorption ionization mass spectrometry-based analysis of wine as a powerful tool for classification purposes

J.D. Nunes-Miranda[a,b], Hugo M. Santos[c,d,e], Miguel Reboiro-Jato[f], Florentino Fdez-Riverola[f], G. Igrejas[b,c], Carlos Lodeiro[a,e], J.L. Capelo[a,e,*]

[a] Bioscope Group, Physical Chemistry Department, Science Faculty, University of Vigo, Ourense, Spain
[b] Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
[c] Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
[d] Unitat d ïEnginyeria de Proteïnes i Proteòmica, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain
[e] REQUIMTE, Departamento de Química, Faculdade de Ciencias e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[f] SING Group, Informatics Department, Higher Technical School of Computer Engineering, University of Vigo, Ourense, Spain

ARTICLE INFO

Article history:
Received 9 November 2011
Received in revised form 7 January 2012
Accepted 9 January 2012
Available online 14 January 2012

Keywords:
Wine
Fingerprinting
MALDI
Classification

ABSTRACT

The variables affecting the direct matrix assisted laser desorption ionization mass spectrometry-based analysis of wine for classification purposes have been studied. The type of matrix, the number of bottles of wine, the number of technical replicates and the number of spots used for the sample analysis have been carefully assessed to obtain the best classification possible. Ten different algorithms have been assessed as classification tools using the experimental data collected after the analysis of fourteen types of wine. The best matrix was found to be α-Cyano with a sample to matrix ratio of 1:0.75. To correctly classify the wines, profiling a minimum of five bottles per type of wine is suggested, with a minimum of three MALDI spot replicates for each bottle. The best algorithm to classify the wines was found to be Bayes Net.

1. Introduction

Current efforts in wine quality control research rely on the development of fast and simple procedures to classify wines [1–4]. Wine traceability is essential to preserve the identity of unique quality traits against frauds or commercial disputes. Accordingly, wineries invest resources in ensuring their customers that their wines are unique and of high quality.

A rapid method to characterize and to classify wines is the so called wine fingerprinting. Each wine has a characteristic pattern of compounds that makes it unique. Such patterns are obtained from components of the wine, such as volatile compounds, proteins, peptides, or other type of organic molecules, such as tannins [1,5,6]. These patterns are generally known as the wine's fingerprint.

At present the use of volatile compounds is the most popular method to classify wines by fingerprinting them. This method relays in the use of gas chromatography mass spectrometry, GC–MS. GC–MS is time consuming, which severely hampers this technique by not allowing it to be used in large-scale trials. Recently, some researchers have focused on the wine's proteomic content in order to use it as the targeted wine's component from which the wine's fingerprint is obtained [1,2,7,8]. The use of the wine's protein/peptide fraction requires intensive sample handling and it is time consuming. In addition, it is also expensive, if large sets of samples are analyzed.

The use of matrix assisted laser desorption ionization, MALDI-based mass spectrometry in fingerprinting has been extensively used to classify samples such as plasma or serum [9–11]. MALDI can be utilized in the direct mode, in which the sample is directly placed onto the MALDI-target for analysis without any sample treatment. The use of the direct-MALDI analysis has a number of advantages, which makes it an ideal tool for wine fingerprinting. First, the sample treatment required is minimum. The wine needs only to be filtered, then the wine is mixed in the MALDI target with an adequate matrix and the sample is ready for analysis. Despite of its great potential for wine classification, the direct-MALDI analysis has not been extensively used yet [12,13]. In addition, the few works already published on this matter do not address key analytical points regarding the MALDI analysis. Thus, a number of parameters such as the normalization of MALDI spectra, the influence of MALDI matrix, the number of technical replicates and the number of MALDI plate spots replicates are missing in the

* Corresponding author at: Bioscope Group, Physical Chemistry Department, Science Faculty/Facultad de Ciencias, As Lagoas, E-32004 Ourense, Spain.
Tel.: +34 610 835 903; fax: +34 988 387 001.
E-mail address: jlcapelom@uvigo.es (J.L. Capelo).

0039-9140/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2012.01.017

Table 1
Selected wines to mass spectrometry analysis.

Code	Wine	Grape Type
A[a]	VegaVerde	Airén, Macabeo
B[a]	Lambrusco Dell'Emilia	Lambrusco
C[a]	L'Antigón	Macabeo, Merseguera
D[a]	Viña do Val	Macabeo, Palomino, Sauvignon Blanc
E[a]	Comportilho Rioja	Viura
F	Coto de Gomariz	Albariño, Godello, Loureira, Treixadura
G	Vilerma Blanco	Albariño, Godello, Treixadura
H	Beade Primacia	Treixadura
I	Gran Reboreda	Treixadura
J	Viña Reboreda	Godello, Palomino, Treixadura, Torrontés
K	Condes de Albarei	Albariño
L	Castillo de Liria	Sauvignon, Viura
M	Pazo Blanco	Treixadura
N	Joaquín Rebolledo	Godello

[a] Represents the wines used for preliminary experiments to determinate the most appropriate MALDI matrix.

academic articles published to date dealing with direct MALDI wine fingerprinting [14].

The present research demonstrates the great potentiality of the direct-MALDI analysis of fingerprinting of wine. Reliable results were obtained after a systematic study of the influence of (i) the type of the MALDI matrix, (ii) the number of the individual, technical and instrumental replicates needed to obtain appropriate data and (iii) the normalization of the intensities of the signals of the MALDI spectra.

2. Experimental

2.1. Reagents

Trifluoroacetic acid (TFA, 99%) was purchased from Riedel-de Haäen (Seelze, Germany). Acetonitrile (ACN, 99.9%) was purchased from Panreac (Barcelona, Spain). All the materials were used without further purification. 2,5-dihydroxybenzoic acid (DHB), sinapinic acid (SA) and alpha-cyano-4-hydroxycinnamic acid(CHCA) for MALDI-TOF-MS were obtained from Fluka (Buchs, Switzerland) and were used as MALDI matrix. Peptide Calibration Standard II from Bruker was used as mass calibration standard for MALDI-TOF-MS.

2.2. Samples

A set of fourteen different Spanish wines presented in Table 1 were selected and five bottles of each wine were acquired from local markets for further analysis.

2.3. Sample treatment

Each sample of wine was filtered through a 0.22 μm pore size cellulose acetate membrane filter. Three MALDI matrices were prepared as follows. Alpha-cyano-4-hydroxycinnamic acid, CHCA, 10 mg/mL in 50% ACN/0.1% TFA. 2,5-dihydroxybenzoic acid, DHB, 20 mg/mL in 90% ACN/0.1% TFA. Sinapinic acid, SA, 20 mg/mL in 30% ACN/0.3% TFA. Each filtered sample was mixed with each matrix in a 1:1 ratio and 1 μL of this mixture was spotted in quintuplicate onto a MALDI-TOF-MS ground steel plate. After that, different ratios between the wines and the MALDI matrix selected were tested in quintuplicate as follows: 1:1; 1:0.75; 1:0.5; 1:0.25.

2.4. MALDI analysis

The MALDI-TOF analysis was performed using the Ultraflex II MALDI-TOF/TOF instrument from Bruker Daltonics equipped with a 200 Hz Smartbeam laser system. Data was acquired using FlexControl 3.3.92.0 (Bruker Daltonics). Close external calibration was performed with the monoisotopic peaks of the bradykinin 1-7 (757.3992), angiotensin II (1046.5418), angiotensin I (1296.6848), substance P (1347.7345), bombesin (1619.8223), renin substrate (1758.9326), ACTH clip 1-17 (2093.0862), ACTH 18-39 (2465.1983), and somatostatin 28 (3147.4710). The mass spectrometer was operated with positive polarity in reflectron mode and spectra were acquired at each spot position at a constant power and in the mass range of 40–1500 Dalton, Da. Peak lists and spectral processing were done in FlexAnalysis 3.3 (Bruker Daltonics). The peak lists were generated from de mass spectra using the peak detection algorithm SNAP (sophisticated numerical annotation procedure). The signal to noise was established at six and the baseline subtraction was done using the TopHat algorithm.

2.5. Machine learning algorithms and computational protocol

In order to systematically study the influence of the analyzed conditions (type of the MALDI matrix, number of the individual and technical replicates and use of the MALDI spectra normalization of signal intensities) the following six well-known Artificial Intelligence, AI, algorithms were selected: Bayes Net [15], C4.5 [16], IBk (*Instance Based k-nearest-neighbor*) [17], Naïve Bayes [18], Random Forest [19] and Support Vector Machine trained with SMO (*Sequential Minimal Optimization*) [20]. In addition, two classical ensemble alternatives, namely AdaBoost [21] and Bagging [22] using different combinations of successful base classifiers, were also used. All the selected classifiers are included in the Weka ML (*Machine Learning*) library [23] and they were executed with default parameters, except for IBk, configured to assess the adequate performance of the proposed models and also to guarantee the validity of the results, we conducted an *ad hoc* LOO (Leave-One-Out) cross-validation

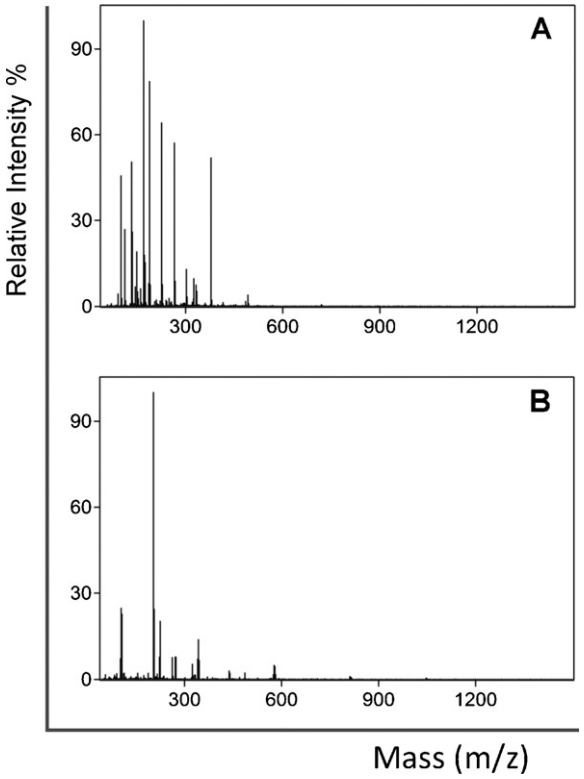

Fig. 1. MALDI-TOF-MS spectra obtained of two different matrices. (A) Mixture of wine and CHCA matrix with a ratio of 1:1. (B) Mixture of wine and sinapinic acid matrix with a ratio of 1:1.

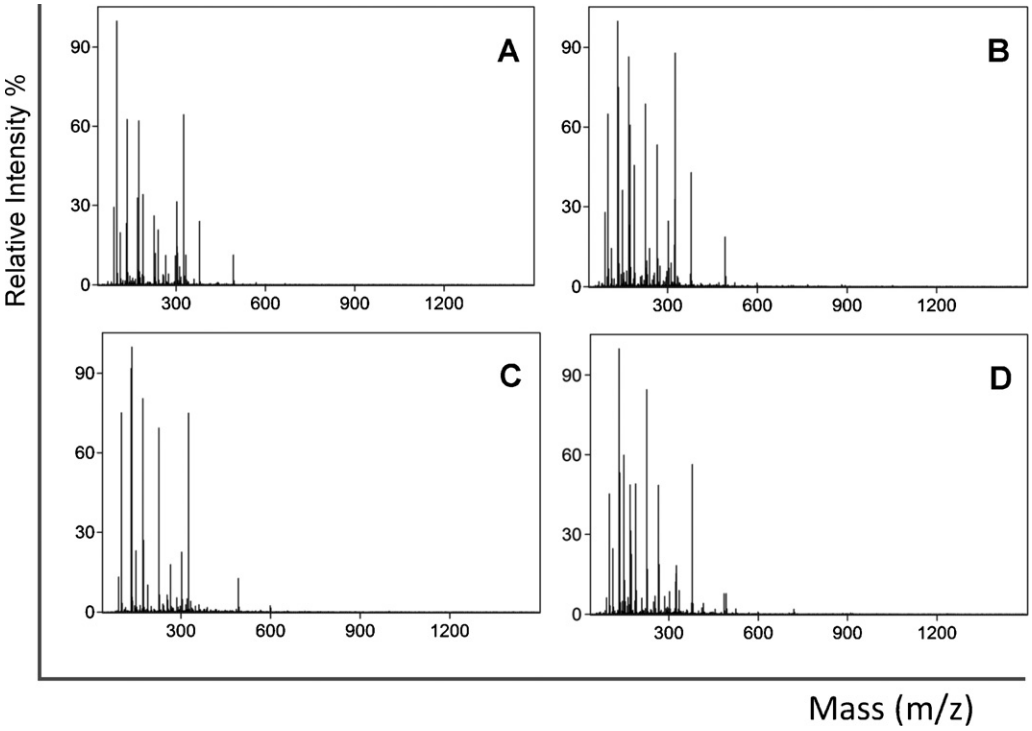

Fig. 2. MALDI-TOF-MS spectra obtained of different ratios of wine/CHCA matrix to infer the best ratio for wine classification. (A) 1:0.25. (B) 1:0.5. (C) 1:0.75. (D) 1:1.

experiment [24] in which each of the 5 spots/bottle were separated form the training samples to predict their corresponding class. All the runs were executed on the AIBench platform [25] using a Weka plugin.

In order to precisely measure the accuracy of each AI technique we usedboth (*i*) the percentage of correct classifications (accuracy) and (*ii*) the Cohen's Kappa statistic [26]. The Kappa index compensates for classifications that may be due to chance and it is considered a standard statistically robust meter useful to measure the accuracy in multiclass problems [27]. Kappa values range from 0 (random classification) to 1 (perfect classification) and it is calculated as shown in Expression (1).

$$Kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (1)$$

where Pr(a) represents the observed accuracy and Pr(e) stands for the probability which is due to chance.

3. Results and discussion

3.1. Influence of the matrix used

It is well known that there is no universal matrix for the MALDI analysis. As a general rule matrixes are chosen experimentally using the trial by error method. From the common matrices generally used in the MALDI analysis we assessed 2,5-dihydroxybenzoic acid, DHB; sinapinic acid and α-Cyano-4-hydroxycinnamic acid, CHCA. The first matrix is best suited for protein, peptides carbohydrates and synthetic polymers, the second one for proteins and peptides and the third one is recommended for peptides. To the best of our knowledge, the analysis of red wine by direct MALDI was only attempted once by Carpenteri et al. [13]. These authors have found that DHB is the most suited MALDI matrix from a series of tested MALDI matrices However, in our laboratory conditions this matrix was found troublesome, since it was verified that some samples took as long as 24 h to dry. This phenomenon was observed for different wines. Therefore, DHB was discharged and further experiments were carried out with sinapinic acid and CHCA.

Fig. 1 shows MALDI spectra of white wine spotted with the aforementioned matrices. A quick inspection of this figure with the naked eye reveals that the spectrum obtained with the α-Cyano has a better signal to noise ratio for most of the peaks than the spectrum corresponding to sinapinic acid. A set of experiments was devised to test the efficiency of both matrices when they were used for fingerprinting wine. The results are shown in Table 2. The comparison was performed using ten different algorithms of classification. In this set of experiments, for each type of white wine (five wines), five bottles were used. For each bottle, one sample was prepared and was spotted in the MALDI plate in quintuplicate. Finally, the spectra thus obtained for each type of wine were used to classify the group of five wines under study. A total of 125 spectra were used (5 wines × 5 bottles/wine × 5 spots/bottle). The peaks of the different spectra were aligned using an error of 150 ppm. Results shown in Table 2 clearly demonstrate that wine classification was better achieved when the spotting was carried out with the CHCA matrix rather than with the sinapinic acid matrix, as a higher accuracy and Kappa scores were obtained. In brief, the closest is the accuracy to 100 and kappa to 1 the better is the classification achieved (for details on statistics refers to Section 2.5). This result was consistently obtained with all the algorithms used. This means that all the algorithms suggest CHCA matrix as the best one to be used to classify the wines. This study was done using only the m/z values of the ions observed in the spectra. However, it was assessed that when m/z values of the ions and their relative intensities were used, the values of accuracy and kappa were improved.

3.2. Influence of matrix to sample ratio

Sample treatment must be as reproducible as possible, especially the matrix/sample crystallization process. Indeed, variables such as temperature, handling time, and number of steps involved in the sample treatment have been identified as important

Table 2
Results of the performance of the CHCA and SA matrices studied. Values of accuracy and Kappa as a function of the classification algorithm used.

Algorithm	With intensities				Without intensities			
	Accuracy		Kappa		Accuracy		Kappa	
	CHCA	SA	CHCA	SA	CHCA	SA	CHCA	SA
Bayes Net	95.20%	84.80%	0.94	0.81	93.60%	78.40%	0.92	0.73
C4.5	94.40%	84.00%	0.93	0.80	90.40%	78.40%	0.88	0.73
IBk	88.00%	59.20%	0.85	0.49	91.20%	64.00%	0.89	0.55
Naïve Bayes	88.00%	80.80%	0.85	0.76	93.60%	76.80%	0.92	0.71
Random forest	94.40%	82.40%	0.93	0.78	90.40%	78.40%	0.88	0.73
SMO	92.80%	84.80%	0.91	0.81	91.20%	77.60%	0.89	0.72
AdaBoost.M1 + IBk	89.60%	63.20%	0.87	0.54	88.80%	72.00%	0.86	0.65
AdaBoost.M1 + C4.5	93.60%	88.80%	0.92	0.86	91.20%	78.40%	0.89	0.73
Bagging + IBk	90.40%	61.60%	0.88	0.52	89.60%	67.20%	0.87	0.59
Bagging + C4.5	93.60%	89.60%	0.92	0.87	86.40%	79.20%	0.83	0.74

parameters. These parameters should be carefully considered and controlled to obtain reproducible data in mass spectrometry fingerprinting of complex samples [28]. The excellent reproducibility between replicates obtained in this work with α-Cyano matrix, can be explained through three main reasons. First, the sample handling is very simple: the wine is filtered. There are no more preparation steps. This reduces sample contamination and sample loses. Second, crystallization with α-Cyano matrix produces small crystals, uniformly distributed throughout the spot, with an excellent reproducibility between spots. Third, baseline drift is corrected by fitting an algorithm that describes de curvature of the baseline and subtracts that from the originating baseline thereby getting a zero baseline. This is done using the software provided by the manufacturer of the MALDI used (Flexanalysis 2.4®). In addition, laboratory temperature was always maintained constant at $21 \pm 1\,^{\circ}C$. Therefore, we studied only the influence of the ratio sample:matrix, because this is one of the variables that most influence MALDI spectra. For instance, if the ratio is too high the matrix ions can hide the signals that belongs to the sample ions. On the other hand, if the ratio is too low, the ionization can fail and no signal is recorded. For this reason a set of experiments was devised to find out the best wine/α-Cyano ratio. The ratios assessed were 1:1; 1:0.75; 1:0.5; 1:0.25. The resulting spectra are shown in Fig. 2 and cannot be inspected with the naked eye, as all the spectra seem similar. However, when an algorithm (SNAP) was used to investigate the differences between them, it was found that the ratio 1:0.75 provided the mass list with the highest record of *m*/*z* ions. Consequently, this ratio was used in further experiments.

3.3. Influence of the number of sample replicates, technical replicates and instrumental replicates on the wine classification.

The MALDI analysis has as main drawback the variance caused by the intra-spot and between spot data variability. As explained above, in our case the performance was very promising rendering data with high accuracy and reproducibility. However, the minimum number of bottles of wine and the minimum number of technical replicates necessary to obtain an appropriate wine classification remained unknown. Therefore a set of experiments was devised as follows: fourteen wines were purchased from local markets. Five bottle of each wine were acquired. From each bottle one sample was prepared (see Section 2), and for each sample five spots were prepared in the MALDI plate. Data was treated with different classification algorithms. From each algorithm a *K* value was retrieved. The medium value of *K*, obtained from the ten algorithms used is presented in Fig. 3 as a function of the number of bottles and the number of spots used for each bottle. In addition, the results for the ten algorithms are presented individually in Fig. 1 of the supplementary material. The classification study was

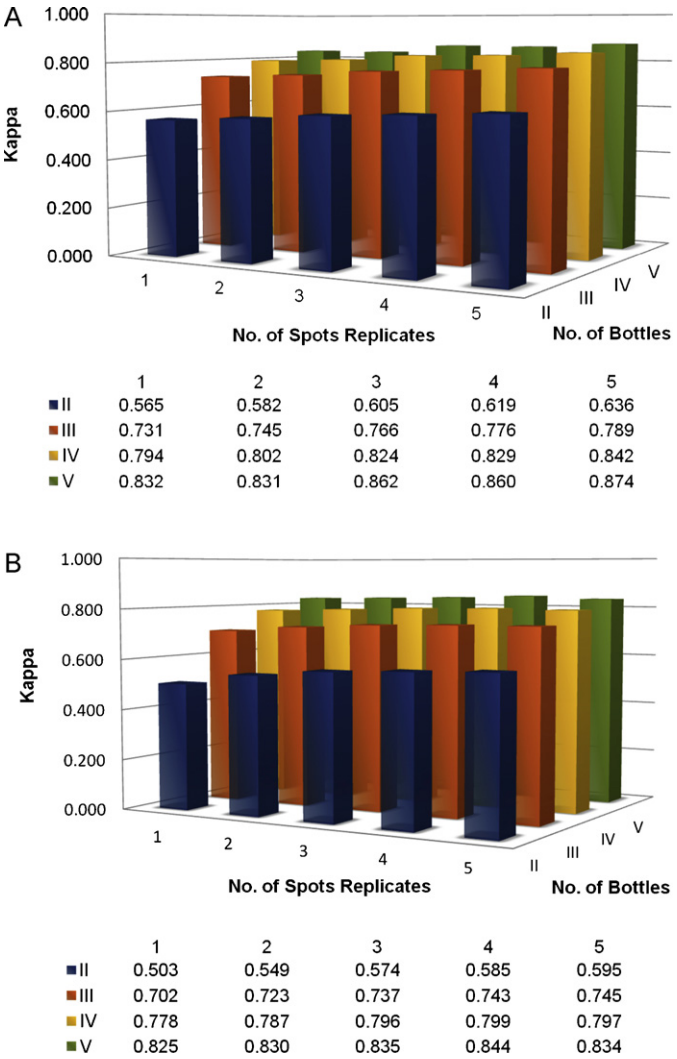

Fig. 3. (A) Average of *K* value obtained for all classifiers used taking into account the influence of the number of replicates and bottles used, with intensities. (B) Average of *K* value obtained for all classifiers used taking into account the influence of the number of replicates and bottles used, without the intensities.

done in two different ways. In the first study, all the information used was the list of *m*/*z* ions. In the second study, *m*/*z* ions and their intensities were used.

It was found that intensity is a variable that adds information, as the classification obtained after adding intensity as a variable

was better. For instance, the value of the K medium was increased from 0.834 to 0.874 for the classification done using five bottles and five spots replicates for each bottle. However, for some of the algorithms the classification was not improved when the intensity was added as additional information. Consequently, it is recommended that researchers routinely repeat their classification with multiple algorithms and consider the *K* medium as the best value.

Data presented in Fig. 3A suggests that the minimum number of spots needed to work with is of three for each wine. More interestingly, the number of bottles to be used seems to be more critical than the number of spots. The minimum number of bottles recommended to be used for classification is five.

4. Future prospects

The results presented here suggest that the MALDI-based fingerprinting of wines is a cheap and fast approach valuable for the classification of the white wine. Through a careful selection of variables such as the number of bottles and the number of MALDI spots replicates used, and through the use of adequate algorithms, wine classification seems to be achievable. The next step in this research will consist in launching a large-scale study, including as many types of grape and wineries as possible. Furthermore, the study will be extended to red wine.

5. Conclusions

It has been demonstrated that the direct MALDI analysis of wine is an effective method to classify wines. A total of fourteen white wines were correctly classified (see supplementary material, Table S1). It has been possible to classify wines by grape type and winery, including three wines made from the same grape but from different wineries. The matrix recommended to be used is α-Cyano with a sample:matrix ratio of 1:0.75. To correctly classify the wines, profiling a minimum of five bottles per type of wine is suggested, with a minimum of three MALDI spot replicates for each bottle. The best algorithm to classify the wines was found to be Bayes Net, although it is recommended to use more than one algorithm to test the robustness of the procedure.

Acknowledgements

Financial support from the Spanish Ministry of Science and Innovation, from the Instituto da Vinha e do Vinho de Portugal and from the University of Vigo under projects ref. AGL2009-09796, ref. 20/2010/SIA and refs. INOU 11B-01 and CO-110-10 respectively, is acknowledged. J. L. Capelo and Carlos Lodeiro, acknowledges the Isidro Parga Pondal research program from the Xunta de Galicia, Spain. H. M. Santos acknowledges the post-doctoral grant SRFH/BPD/73997/2010 provided by the FCT-MCTES (Fundação para a Ciência e a Tecnologia–Ministério da Ciência, Tecnologia e Ensino Superior, Portugal). M. Reboiro-Jato was supported by a pre-doctoral fellowship from University of Vigo. J.D. Nunes-Miranda thanks Scientific Association PROTEOMASS for financial support given.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2012.01.017.

References

[1] A. Chambery, G. del Monaco, A. Di Maro, A. Parente, Food Chem. 113 (2009) 1283–1289.
[2] D. Le Bourse, S. Conreux Jégou, S. Villaume, P. Jeandet, Anal. Chim. Acta 667 (2010) 33–42.
[3] P. Tarantilis, V. Troianou, C. Pappas, Y. Kotseridis, M. Polissiou, Food Chem. 111 (2008) 192–196.
[4] R.R. Catharino, I.B.S. Cunha, A.O. Fogaça, E.M.P. Facco, H.T.C. Godoy, C.E. Daudt, M.N. Eberlin, A.C.H.F. Sawaya, J. Mass Spectrom. 41 (2006) 185–190.
[5] I.C. Pesavento, A. Bertazzo, R. Flamini, A.D. Vedova, M.D. Rosso, R. Seraglia, P. Traldi, J. Mass Spectrom. 43 (2008) 234–241.
[6] R. Flamini, Mass Spectrom. Rev 22 (2003) 218–250.
[7] D. Chabreyrie, S. Chauvet, F. Guyon, H.M. Salagoïty, J.F. Antinelli, B. Medina, J. Agric. Food Chem. 56 (2008) 6785–6790.
[8] S.W. Known, J. Agric, Food Chem. 52 (2004) 7258–7263.
[9] S. Alagaratnam, B.J.A. Mertens, J.C. Dalebout, A.M. Deelder, G.-J.B. van Ommen, J.T. den Dunnen, P.A.C. 't Hoen, Proteomics 8 (8) (2008) 1552–1563.
[10] L. Long, R. Li, Y. Li, C. Hu, Z. Li, Rheumatol. Int. 31 (2011) 1069–1074.
[11] M. Pietrowska, L. Marczak, J. Polanska, K. Behrendt, E. Nowicka, A. Walaszczyk, A. Chmura, R. Deja, M. Stobiecki, A. Polanski, R. Tarnawski, P. Widlak, J. Transl. Med. 7 (60) (2009), doi:10.1186/1479-5876-7-60.
[12] Z. Szilágyi, G. Vas, G. Mády, K. Vékey, Rapid Commun. Mass Spectrom. 10 (1996) 1141–1143.
[13] A. Carpentieri, G. Marino, A. Amoresano, Anal. Bioanal. Chem. 389 (2007) 969–982.
[14] J.D. Nunes-Miranda, G., Igrejas, J.C. Mejuto, M. Reboiro-Jato, J.L. Capelo-Martinez, Crit. Rev. Food. Sci. (2012) Manuscript in press.
[15] J. Pearl Bayesian Networks, Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA., 1985, pp. 329–334.
[16] R. Quinlan, C4. 5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
[17] D. Aha, D. Kibler, Mach. Learn. 6 (1991) 37–66.
[18] G.H. John, P. Langle, Proceeding of the 11 h Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995, pp. 338–345.
[19] L. Breiman, Mach. Learn. 45 (2001) 5–32.
[20] J. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods – Support Vector Learning, MIT Press, Cambridge, MA, USA, 1998.
[21] Y. Freund, R.E. Schapire, J. Comput. Syst. Sci. 55 (1997) 119–139.
[22] L. Breiman, Mach. Learn. 24 (1996) 123–140.
[23] M. Hall, E. Frank, H. Geoffrey, B. Pfahringer, P. Reutemann, I.H. Witten, SIGKDD Explorations 11 (2009) 10–18.
[24] R. Kohavi, Proceedings of the 14th international joint conference on Artificial intelligence, 1995, Vol. 2, pp 1137 1143.
[25] D. Glez-Peña, M. Reboiro-Jato, P. Maia, M. Rocha, F. Díaz, F. Fdez-Riverola, Comput. Meth. Prog. Biol. 98 (2010) 191–203.
[26] J. Cohen, Educ. Psychol.Meas. 20 (1960) 37–46.
[27] A. Ben-David, Expert Sys. Appl. 34 (2008) 825–832.
[28] M. West-Norager, C.D. Kelstrup, C. Schou, E.V. Hogdall, C.K.N.H. Hogdall, H. Heegaard, J. Chromatogr. B847 (2007) 30–37.